Modern Classical Optics

GEOFFREY BROOKER

Department of Physics
University of Oxford

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Published in the United States
by Oxford University Press Inc., New York

First published 2003

A catalogue record for this title
is available from the British Library

Library of Congress Cataloging in Publication Data
(Data available)

ISBN 0 19 859965 X (Pbk)
ISBN 0 19 859964 1 (Hbk)

10 9 8 7 6 5 4 3 2 1

Typeset by the author using LaTeX

Printed in Great Britain
on acid-free paper by The Bath Press, Avon

The Fabry–Perot

5

5.1 Introduction

This chapter examines a particular case of multiple-beam interference: where light is bounced repeatedly between two highly reflecting mirrors. It is assumed that the reader has already encountered interference 'by division of amplitude' in its two-beam case, as happens when light is reflected from opposite faces of a slab or traces the two paths through a Michelson interferometer. It is further assumed that the following are familiar:[1]

- The optical path difference is $n\,2d\cos\theta$ when a light beam divides at the surfaces (separation d) of a slab of refractive index n, and travels inside the slab at angle θ to the normal.
- In the case where the reflecting surfaces are parallel, interference fringes 'of equal inclination' are localized at infinity.

We have already encountered a case where multiple beams of light are interfered: those beams that pass through the many elements of a diffraction grating and are combined 'downstream'.[2] In both cases, the advantage of 'multiple' is a sharpening of the fringes, with a consequent enhancement of resolution, according to the value of a quantity called the finesse.

The Fabry–Perot is worth understanding in its own right as a spectroscopic instrument. It is also a useful model to have in mind when understanding thin films (Chapter 6) and laser cavities (Chapter 8).

[1] An unusual derivation of the optical path difference is outlined in problem 5.1, and fringe localization is explained in §11.11.

[2] Interference at a grating can be made to resemble that with a Fabry–Perot, as is shown in problem 5.11.

5.2 Elementary theory

Figure 5.1 shows a light ray (meaning a wave whose wavefronts are normal to the given ray) arriving from bottom left and incident on the first of two identical partly reflecting mirrors. Some light passes through into the space between the mirrors. There it is repeatedly reflected, with a small fraction transmitted at each encounter with a mirror.

At the first surface, the incident wave of amplitude U_0 has a fraction t_1 of its amplitude transmitted. At the second surface, a fraction t_2 is transmitted and a fraction r_2 is reflected. After each round trip, the light that remains between the mirrors has undergone two reflections (amplitude reflection coefficient r_2) and a phase lag of $k\,2d\cos\theta$ resulting from travel through distance $2d\cos\theta$. The outcome is a set of transmitted wave amplitudes, the first few of which are marked on Fig. 5.1.

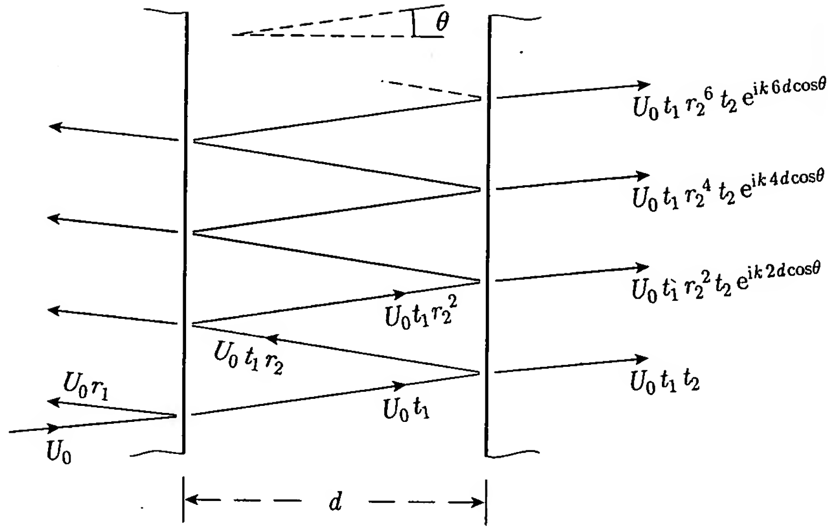

Fig. 5.1 The path of a representative wave through a Fabry–Perot. The wave is partially reflected and partially transmitted at each of the two reflecting surfaces. Amplitude transmission coefficients t_1, t_2 are defined on the diagram, as are the 'outside' amplitude reflection coefficient r_1 and the 'inside' reflection coefficient r_2. The commonest configuration has an air gap between two spaced glass blocks, and such a structure is hinted at in the drawing.

The coefficients t_1, t_2, r_1, r_2 may be complex if there are phase shifts at the mirrors, as there may well be. Let the 'internal' intensity reflection coefficient be $R = |r_2|^2$ and let $r_2^2 = R\,e^{-i\alpha}$. The total phase lag between successive transmitted beams is now $\delta = k\,2d\cos\theta - \alpha$, where this allows for the round-trip travel and the effect of two reflections. (The reflection phase shift α is included for honesty only and will sometimes be ignored to avoid cluttering the discussion.)

The total amplitude transmitted through the Fabry–Perot is now

$$U_{\text{out}} = U_0\, t_1 t_2 \{1 + R\,e^{i\delta} + R^2\,e^{2i\delta} + \cdots\} = \frac{U_0 t_1 t_2}{1 - R\,e^{i\delta}}, \tag{5.1}$$

where we use the sum of a geometric progression.

The transmitted intensity $I_{\text{transmitted}}$ evaluates to

$$\frac{I_{\text{transmitted}}}{I_{\text{incident}}} = \frac{|U_{\text{out}}|^2}{|U_0|^2} = \left(\frac{|t_1 t_2|}{1-R}\right)^2 \frac{1}{1 + \dfrac{4R}{(1-R)^2}\sin^2(\delta/2)}. \tag{5.2}$$

The prefactor $|t_1 t_2|/(1-R)$ is of little interest: it varies only slowly with θ (depending on the physics of the reflecting surfaces) and not at all with d, so it will henceforth be absorbed into I_{max}, the maximum transmitted intensity. Summarizing and tidying a little, we now have:

$$\frac{I_{\text{trans}}}{I_{\text{max}}} = \frac{1}{1 + (4\mathcal{F}^2/\pi^2)\sin^2(\delta/2)} \qquad \text{the Airy function} \tag{5.3}$$

$$\mathcal{F} = \pi\sqrt{R}/(1-R) \qquad \text{the finesse} \tag{5.4}$$

$$\delta = k\,2d\cos\theta - \alpha = (\omega/c)n\,2d\cos\theta - \alpha \qquad \text{the phase lag.} \tag{5.5}$$

Here n is the refractive index of the medium between the reflecting surfaces, and θ is the angle between the rays and the normal *between the*

reflectors. Some straightforward mathematical steps have been omitted in the above, and are rehearsed in problem 5.2.

Comment 1 The single incident ray represents, as usual, a broad wavefront arriving from some source, details of which are not shown. Since only one ray is drawn, we are not told whether the incident wavefront is plane or spherical or of some other shape; we shall see that it does not matter. Because the incident wave is broad, the transmitted waves are broad also, and can be taken to overlap each other even though their rays are drawn with a lateral displacement.[3]

Comment 2 Fabry–Perot devices[4] are most commonly made from two glass blocks, with the reflecting surfaces facing 'inwards' to an air gap. The amplitude transmission coefficients t_1 and t_2 are necessarily unequal, because the light in one case travels from glass to air and in the other from air to glass.

Comment 3 The finesse $\mathcal{F}$ has been constructed from the group of factors $4R/(1-R)^2$ in a way that may not seem obvious. We shall see that the finesse $\mathcal{F}$, as defined, is the most useful quantity for specifying the quality of the interference fringes.[5]

Comment 4 No attempt has been made here at simplifying or evaluating the prefactor $|t_1t_2|/(1-R)$. Since the reflecting surfaces might be as different as evaporated aluminium or several layers of dielectric, no general simpler expression could exist.[6]

Comment 5 Although the calculation leading to eqns (5.3)–(5.5) is in every textbook, we should not accept it uncritically as the only possible way forward. Problem 5.3 offers a rather different analysis. Problem 5.4 asks whether eqn (5.3) meets tests for reasonableness.

5.3 Basic apparatus

The addition of amplitudes that was performed in obtaining eqn (5.1) represents, of course, the interference of the successively reflected waves, each differing in phase from the one before. Equation (5.5) shows that the phase difference varies with θ only, if d and n are held fixed. Whether we get light or dark then depends upon θ.

In order to display the interference, we need to gather all light of a given θ and put it into its own place, separate from all other θs. Figure 1.3 explains how this may be done, because a lens focused for infinity selects angles in just the way we need. A minimal arrangement of apparatus is, therefore, that shown in Fig. 5.2. In the focal plane of the lens there are fringes in the form of bright circular rings,[7] narrow if the finesse $\mathcal{F}$ is large.

Since the optics has its correct geometry only when we look in the focal plane of the fringe-forming lens, we say that the fringes are *localized at infinity.* (For more on fringe localization, see §11.11.)

By contrast, restrictions on the 'upstream' side of the Fabry–Perot are conspicuously absent. The single input ray drawn in Fig. 5.1 is

[3]Under the conditions envisaged in this comment, interference takes place everywhere that there is light on the downstream side of the Fabry–Perot. If the interference is constructive, light is transmitted; if not, it is reflected. The lens introduced in §5.3 is not required to make the interference happen, but to unscramble what would otherwise be a confusing mess.

[4]A pair of reflecting plates mounted either side of a fixed-width spacer constitutes a Fabry–Perot **étalon**. An arrangement in which one plate may be moved on rails to vary d is called a Fabry–Perot **interferometer**. In this book we write 'Fabry–Perot' when (as is usually the case) the discussion applies to devices of either construction; we refer to an 'étalon' only when concerned specifically with a fixed-d device.

The French word 'étalon' has two meanings: a standard, and a stallion. It is usually clear which is to be found in a laboratory.

[5]Some books write $F = 4R/(1-R)^2$ and work with this F as well as (or instead of) $\mathcal{F}$. Here we avoid F in an effort at minimizing confusion.

[6]There is a theorem, investigated in problem 6.5(4), that $|t_1t_2|/(1-R) = 1$ for the special case of a loss-free reflecting surface. This property applies even if the 'surface' is composite so long as it is loss-free. Our grounds for discarding the prefactor as uninteresting were not, however, based on any assumption that the reflectors might be loss-free.

[7]The fringe pattern must be circular because the optical system has axial symmetry about an axis normal to the étalon plates and passing through the centre of the lens.

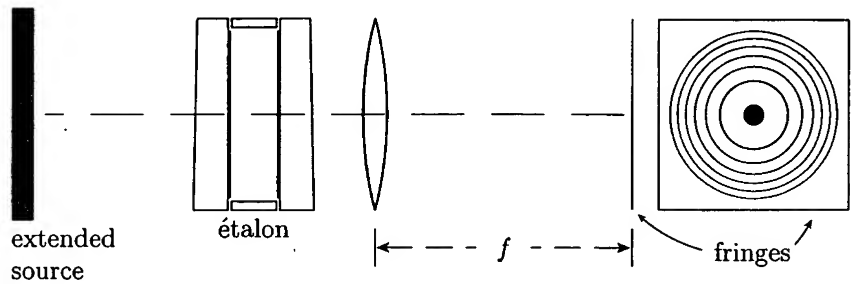

Fig. 5.2 A 'minimal' apparatus for exhibiting Fabry–Perot fringes. The étalon plates and spacer are shown, but the mount for holding them together is omitted to avoid clutter. The 'fringe-forming' lens selects light travelling in each direction θ and puts it in its own place in the focal plane; it maps directions onto positions.

consistent with any wavefront geometry in the input light. And Fig. 5.2 shows a broad illuminating source with no 'preparation' of any kind of the light from that source. There are in fact usually good reasons for elaborating the apparatus of Fig. 5.2, but those reasons are secondary, such as optimizing the brightness of the fringes, rather than getting fringes to happen at all. Such matters are discussed in problem 5.5. See also sidenote 24 on p. 122.

5.4 The meaning of finesse

The intensity transmitted by a Fabry–Perot, given in eqn (5.3), has peaks that occur where $\sin(\delta/2) = 0$, which means that $\delta = p\,2\pi$, where p is an integer. Equivalently, and ignoring α for simplicity, there are

$$\text{maxima of transmitted intensity where} \quad n\,2d\cos\theta = p\lambda_{\text{vacuum}}. \tag{5.6}$$

Figure 5.3 shows $I_{\text{trans}}/I_{\text{max}}$ plotted versus the phase difference δ, covering a range of δ that includes the maxima of orders p to $p+1$.

We shall find the FWHM, the full width at half maximum height, of the pth peak by finding where $I_{\text{trans}}/I_{\text{max}} = \frac{1}{2}$. From eqn (5.3), this happens at $\delta = p\,2\pi + \varepsilon$ where

$$\pm\pi/(2\mathcal{F}) = \sin(\delta/2) = (-1)^p \sin(\varepsilon/2),$$

so that

$$\varepsilon = 2\arcsin(\pm\pi/2\mathcal{F}).$$

In the cases of interest, $\mathcal{F}$ is made to be large (perhaps 30 or more, achieved by using high reflectivity R), so ε is small[8] (though δ is not), and ε can be evaluated by using the power series for arcsin. The result is: $\varepsilon = \pm\pi/\mathcal{F} + \text{order } \mathcal{F}^{-3}$. The FWHM of the peak, measured along the δ axis, is thus $2\pi/\mathcal{F}$ and

$$\frac{\text{separation of peaks}}{\text{FHWM of one peak}} \approx \frac{2\pi}{2\pi/\mathcal{F}} = \mathcal{F}. \tag{5.7}$$

This result explains why the finesse $\mathcal{F}$ is the useful measure of Fabry–Perot fringe quality.

[8]Phase shift δ is *not* small because the integer p may well be of the order of thousands.

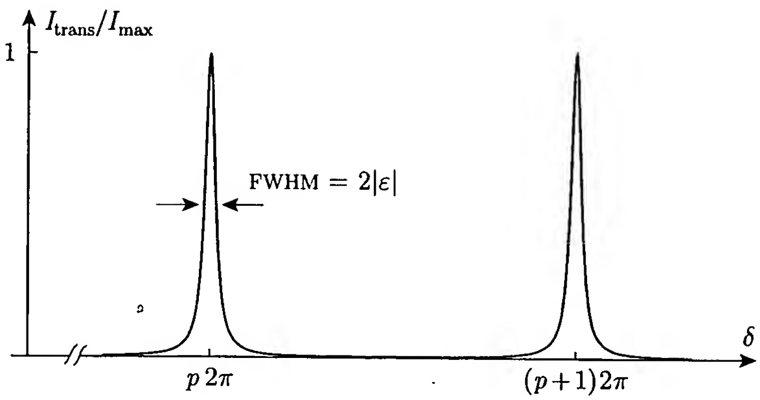

Fig. 5.3 A graph of $I_{\text{trans}}/I_{\max}$, the normalized intensity transmitted, plotted against phase shift δ. The graph is drawn for a finesse of 20, a modest value, but one that helps with displaying the details of the curve.

5.5 Free spectral range and resolution

The main use for a Fabry–Perot is as a spectroscopic device for investigating a group of closely spaced spectral lines.

Figure 5.4 shows a graph of the transmitted intensity I_{trans} plotted against $n2d\cos\theta$ (not δ this time, note).[9] We apply two slightly different frequencies ν, ν' (vacuum wavelengths λ, λ', wavenumbers $\bar{\nu}$, $\bar{\nu}'$). Frequency ν forms a system of rings; ν' forms a second system of rings, interlacing those for ν. Our aim is to understand how measurements can reveal that two distinct frequencies are present and yield a value for the frequency difference $(\nu' - \nu)$.

[9]The graph of Fig. 5.4 can be understood as describing families of rings if we regard $\cos\theta$ as varying while (nd) is held fixed. But the abscissa has been chosen so that the graph is equally applicable to cases where one of the other variables is being scanned. We shall see that in practice it more usual to scan n while holding $d\cos\theta$ fixed.

Figure 5.7 shows a plot that is the result of scanning the refractive index n.

5.5.1 Free spectral range

Imagine that ν' starts equal to ν, and that we are able gradually to tune ν' while watching the two ring families as they separate. To keep free from confusion, we want the peak for $p\lambda'$ to stay between $p\lambda$ (where it started) and $(p+1)\lambda$. The largest permitted change brings $p\lambda'$ just up

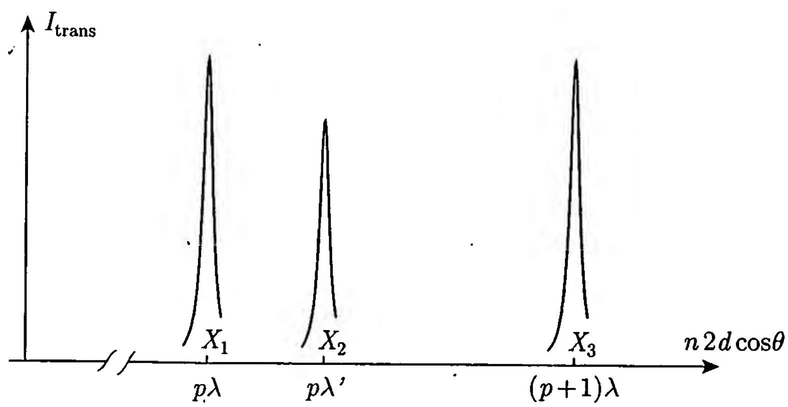

Fig. 5.4 The intensity I_{trans} transmitted by a Fabry–Perot, drawn as a function of $(n\,2d\cos\theta)$, when two wavelengths λ, λ' are input. Wavelength λ' is imagined to be adjustable so that point X_2 can be moved along the graph. To avoid confusion, X_2 should not normally be allowed to reach or pass X_3, a condition that fixes the free spectral range.

to $(p+1)\lambda$, at which point $\lambda' = \lambda'_F$, $\bar{\nu}' = \bar{\nu}'_F$, with

$$p\lambda'_F = (p+1)\lambda, \qquad p/\bar{\nu}'_F = (p+1)/\bar{\nu}, \qquad p(\bar{\nu} - \bar{\nu}'_F) = \bar{\nu}'_F = 1/\lambda'_F.$$

The **free spectral range** is this largest permitted change of frequency, quantified in terms of wavenumber $\bar{\nu} = 1/\lambda_{\text{vac}} = \nu/c$:

$$\text{free spectral range} = \bar{\nu} - \bar{\nu}'_F = \frac{1}{p\lambda'_F} = \frac{1}{n\,2d\cos\theta} \approx \frac{1}{2d}. \tag{5.8}$$

The approximation in the final step is permitted only now, after nearly equal quantities have been subtracted.

5.5.2 Resolution

Return to Fig. 5.4. Imagine tuning ν' starting from ν, but this time moving the $p\lambda'$ peak only just enough to separate it from $p\lambda$ by the width (FWHM) of one peak. We take this to be the setting at which the two frequencies are just resolved.[10] Equation (5.7) tells us that the FWHM is $1/\mathcal{F}$ of the separation between peaks (adjacent, belonging to the same wavelength). So $p\lambda'$ has been tuned by $1/\mathcal{F}$ of the distance from $p\lambda$ to $(p+1)\lambda$, that is, by $1/\mathcal{F}$ of free spectral range:

$$\text{least resolvable wavenumber difference } \delta\bar{\nu} = \frac{\text{free spectral range}}{\text{finesse}}. \tag{5.9}$$

Problem 5.6 draws attention to some subtleties that entered into the correct construction of the reasoning above.

The 'least resolvable wavenumber difference' is more commonly referred to as the *instrumental width*, meaning the apparent width in frequency of monochromatic light, as signalled by the FWHM of the peaks in the output from the Fabry–Perot.[11] Equation (5.9) can be recast as

$$\text{instrumental width} = \delta\bar{\nu} = \frac{1}{n\,(2d) \times (\text{finesse})}. \tag{5.10}$$

Comparing this with eqn (4.6), we see that[12]

$$n\,(2d) \times (\text{finesse}) = \begin{pmatrix}\text{effective longest optical path} \\ \text{difference for interference}\end{pmatrix}. \tag{5.11}$$

For this reason, the finesse is often thought of as the 'effective' number of round trips undertaken by the light as it bounces between the two reflecting surfaces.

Finally, we return to eqn (5.7) and Fig. 5.4. Two adjacent peaks belonging to a single frequency are separated (with our choice of abscissa) by λ. The width (FWHM) of a peak is $1/\mathcal{F}$ of this. So the number of additional peaks (belonging to other frequencies) that could just be fitted edge-to-edge[13] into one order is $\mathcal{F}$. The finesse $\mathcal{F}$, therefore, gives us a measure of the complexity of a spectrum which the Fabry–Perot could be expected to display, given an ideal choice of free spectral range.

Problems 5.8–5.10 investigate the resolution and the instrumental width from several points of view that complement that given here.

[10] This choice is sometimes called the Taylor criterion, in contrast to the Rayleigh criterion used in connection with a diffraction grating.

[11] Instrumental width has already been met in connection with a grating in eqn (4.8) and problem 4.7, and the idea is further developed in problem 5.8.

[12] As with eqn (4.6), the accuracy we need when specifying resolution does not usually justify our bothering with the refractive index in eqn (5.11).

[13] We can't, of course, expect to separate and analyse peaks if they really are fitted in so tightly. A safety factor of perhaps 3–5 is needed. But the finesse correctly gives us the estimate to which such safety factors should be applied.